Econometric Analysis

Ryuichi Tanaka

Ch. 10 Basic Linear Unobserved Effects Panel Data Models

10.1 Motivation: The Omitted Variables Problem

$(y, x_1, \cdots, x_K)$: observable random variables

(c): unobserved effect (ex. cognitive ability, motivation)

$E(y \mid x_1, \cdots, x_K, c) = \beta_0 + \mathbf{x}\boldsymbol{\beta} + c$

If $Cov(x_j, c) \neq 0$**,OLS for** β **is inconsistent**

<u>**Three consistent estimators for** β</u>

（１） proxy variable

（２） IV (2SLS)

（３） Multiple indicator IV

Other estimators when panel data are available

— random effect, fixed effect, first differencing

$(y_t, \mathbf{x}_t)$, $t = 1,2$, c: constant across time

$E(y_t \mid \mathbf{x}_t, c) = \beta_0 + \mathbf{x}_t\boldsymbol{\beta} + c$, $t = 1,2$

Error form: $y_t = \beta_0 + \mathbf{x}_t\boldsymbol{\beta} + c + u_t$, where $E(u_t \mid \mathbf{x}_t, c) = 0$, $t = 1,2$ (=> $E(\mathbf{x}_t' u_t) = \mathbf{0}$, $t = 1,2$)

If $E(\mathbf{x}_t' c) \neq \mathbf{0}$**,** pooled OLS is biased and inconsistent!

<u>First differencing</u>: $\Delta y = \Delta\mathbf{x}\boldsymbol{\beta} + \Delta u$, where $\Delta y = y_2 - y_1$ $\Delta\mathbf{x} = \mathbf{x}_2 - \mathbf{x}_1$ $\Delta u = u_2 - u_1$

Under AOLS1 ($E(\Delta\mathbf{x}'\Delta u) = \mathbf{0}$) and **AOLS2 (**$rank E(\Delta\mathbf{x}'\Delta\mathbf{x}) = K$**), OLS is** consistent.

<u>**On AOLS1**</u>

$E(\Delta\mathbf{x}'\Delta u) = \mathbf{0}$ <=> $E(\mathbf{x}_1' u_1) + E(\mathbf{x}_2' u_2) - E(\mathbf{x}_2' u_1) - E(\mathbf{x}_1' u_2) = 0$

$E(\mathbf{x}_t' u_t) = \mathbf{0}$ **(contemporaneous exogeneity)** => $E(\mathbf{x}_1' u_1) = E(\mathbf{x}_2' u_2) = 0$

strict exogeneity => $E(\mathbf{x}_1' u_1) = E(\mathbf{x}_2' u_2) = E(\mathbf{x}_2' u_1) = E(\mathbf{x}_1' u_2) = 0$

<u>**On AOLS2**</u>

If $x_{j1} = x_{j2}$**, the j-th column of the matrix** $\Delta\mathbf{x} = \mathbf{x}_2 - \mathbf{x}_1$ **becomes 0 (full-rank condition AOLS2 is not satisfied)**

Variation across time in $\mathbf{x}_{it}$ **is needed for consistent estimation of beta**

10.2.1 Random or Fixed Effects?

The unobserved effect model (UEM)

$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}$, $t = 1,\cdots,T$, c_i : unobserved component, unobserved heterogeneity

Some call c_i random effect if it is a random variable, and fixed effect is it is a parameter.

However, it is natural to consider c_i as a random variable drawn from the population along with $(y_{it}, \mathbf{x}_{it})$ (by Wooldridge)

Definition (in this lecture)

Random effect: $Cov(\mathbf{x}_{it}, c_i) = 0$, $t = 1,\cdots,T$

Fixed effect: $Cov(\mathbf{x}_{it}, c_i) \neq 0$ for some $t = 1,\cdots,T$.

10.2.2 Strict Exogeneity Assumptions on the Explanatory Variables

$(\mathbf{x}_{i1},\cdots,\mathbf{x}_{iT})$ is **strictly exogenous** **conditional on the unobserved effect** c_i if

$E(y_{it} \mid \mathbf{x}_{i1},\cdots,\mathbf{x}_{iT}, c_i) = E(y_{it} \mid \mathbf{x}_{it}, c_i) = \mathbf{x}_{it}\boldsymbol{\beta} + c_i$, $t = 1,\cdots,T$.

Controlling $(\mathbf{x}_{it}, c_i)$, other $\mathbf{x}_{is}$, $t \neq s$ has no partial effect on y_{it}

Another expression of strict exo. for $y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}$: $E(u_{it} \mid \mathbf{x}_{i1},\cdots,\mathbf{x}_{iT}, c_i) = 0$, $t = 1,\cdots,T$

$E(u_{it} \mid \mathbf{x}_{i1},\cdots,\mathbf{x}_{iT}, c_i) = 0$ => $E(\mathbf{x}_{is}'u_{it}) = \mathbf{0}$ for $s,t = 1,\cdots,T$

The assumption $E(\mathbf{x}_{is}'u_{it}) = \mathbf{0}$ for $s,t = 1,\cdots,T$ is enough for consistency, but $E(u_{it} \mid \mathbf{x}_{i1},\cdots,\mathbf{x}_{iT}, c_i) = 0$ is needed for statistical inference

10.2.3 Some Examples of Unobserved Effects Panel Data Models

Two issues for panel analysis

1. Is unobserved effect c_i **correlated to other observed explanatory variables $\mathbf{x}_{it}$? (if yes, the effect is fixed effect. If no, random effect)**

2. Is the assumption on $(\mathbf{x}_{i1},\cdots,\mathbf{x}_{iT})$ of **strictly exogenous conditional on the unobserved effect c_i reasonable? (If yes, estimable consistently. If no, need to consider another estimator)**

Example 10.1: Program eveluation

$\log(wage_{it}) = \theta_t + \mathbf{z}_{it}\boldsymbol{\gamma} + \delta_1 prog_{it} + c_i + u_{it}$

Since program participation may depend on unobserved characteristics such as motivation and ability, include c_i in the population model

1. Is unobserved effect c_i correlated to observed explanatory variables $\mathbf{x}_{it}$?

Ex. participants ($prog_{it} = 1$) have high unobserved ability (c_i)(self-selection problem)

In this case, we need to consider a fixed effect model

2. Is the assumption of **strictly exogenous conditional on the unobserved effect** reasonable?

Ex. If only workers with low wage (low u_{it}) are eligible for the next period's program ($prog_{it+1}=1$), strict exogeneity is not satisfied

If a model contains lagged dependent variable as an explanatory variable, strict exogeneity is not satisfied. Moreover, since lagged dependent variables must correlate to unobserved effect c_i, we cannot use a random effect model

10.3 Estimating Unobserved Effects Models by Pooled OLS

The unobserved effect model (UEM): $y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}$, $t = 1,\cdots,T$

<u>**Necessary condition for Pooled OLS to be consistent:**</u> $E(\mathbf{x}_{it}'(c_i + u_{it})) = 0$.

=> $E(\mathbf{x}_{it}'u_{it}) = \mathbf{0}$ and $E(\mathbf{x}_{it}'c_i) = \mathbf{0}$

$E(\mathbf{x}_{it}'c_i) = \mathbf{0}$ **is a strong assumption** $E(\mathbf{x}_{it}'c_i) = \mathbf{0}$ is not satisfied in a distributed lag model

Even if these conditions are satisfied, we need to use robust variance matrix for inference because error term of OLS $v_{it} = c_i + u_{it}$ **is serially correlated**

10.4 Random Effects Methods (RE is **FGLS with special form of** $\boldsymbol{\Omega} = E(\mathbf{v}_i\mathbf{v}_i')$)

10.4.1 Estimation and Inference under the Basic Random Effects Assumptions

The unobserved effect model (UEM): $y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}$, $t = 1,\cdots,T$

<u>**Basic Idea**</u> : Under the assumption that **composite error** $v_{it} = c_i + u_{it}$ **does not depend on explanatory variable(**$E(v_{it} \mid \mathbf{x}_i) = 0$, $t = 1,\cdots,T$**), do GLS taking into account the serial correlation in** $v_{it} = c_i + u_{it}$

<u>**Assumption RE.1**</u>: (a) $E(u_{it} \mid \mathbf{x}_i, c_i) = 0$, $t = 1,\cdots,T$ (strict exogeneity)

(b) $E(c_i \mid \mathbf{x}_i) = E(c_i) = 0$ (orthogonality between c and x)

For each i, stack T observations

The unobserved effect model: $\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{v}_i$

variance matrix: $\boldsymbol{\Omega} = E(\mathbf{v}_i\mathbf{v}_i')$

<u>**Assumption RE.2**</u>: $rank E(\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\mathbf{X}_i) = K$

Under ARE1 and ARE. 2, FGLS (RE) estimation is consistent.

Assumption RE.3: (a) $E(\mathbf{u_i u_i}' | \mathbf{x}_i, c_i) = \sigma_u^2 \mathbf{I}_T$, (b) $E(c_i^2 | \mathbf{x}_i) = \sigma_c^2$

Under ARE.3, $\mathbf{\Omega} = E(\mathbf{v_i v_i}')$ has the **random effects structure:**

$$\mathbf{\Omega} = E(\mathbf{v_i v_i}') = \begin{pmatrix} \sigma_c^2 + \sigma_u^2 & \sigma_c^2 & \cdots & \sigma_c^2 \\ \sigma_c^2 & \sigma_c^2 + \sigma_u^2 & \cdots & \vdots \\ \vdots & & \ddots & \sigma_c^2 \\ \sigma_c^2 & & & \sigma_c^2 + \sigma_u^2 \end{pmatrix}$$ (Error term correlates only through c)

FGLS with consistent estimators for σ_c^2, σ_u^2

$\hat{\sigma}_v^2 = \hat{\sigma}_c^2 + \hat{\sigma}_u^2 = \frac{1}{NT-K}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{v}_{it}^2$, $\hat{v}_{it}$: Pooled OLS residual

$\hat{\sigma}_c^2 = \frac{1}{NT(T-1)/2-K}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\hat{v}_{it}\hat{v}_{is}$ (because $\sigma_c^2 = E(v_{it}v_{is})$, $t \neq s$)

Under ARE1, 2, 3, RE estimator is efficient because it is asymptotically equivalent to GLS

10.4.2 Robust Variance Matrix Estimator (skip)

Without ARE3, inference based on robust variance matrix because RE is consistent under ARE1 and ARE2

asymptotic variance

$\hat{\mathbf{V}} = N^{-1}\hat{\mathbf{A}}^{-1}\hat{\mathbf{B}}\hat{\mathbf{A}}^{-1}$

$\hat{\mathbf{A}} = N^{-1}\sum_{i=1}^{N}\mathbf{X_i}'\hat{\mathbf{\Omega}}^{-1}\mathbf{X_i}$ **(consistent estimator of A)**

$\hat{\mathbf{B}} = N^{-1}\sum_{i=1}^{N}\mathbf{X_i}'\hat{\mathbf{\Omega}}^{-1}\hat{\mathbf{v}}_i\hat{\mathbf{v}}_i'\hat{\mathbf{\Omega}}^{-1}\mathbf{X_i}$, $\hat{\mathbf{v}}_i = \mathbf{y_i} - \mathbf{X_i}\hat{\boldsymbol{\beta}}_{RE}$ **(RE residual)**

10.4.3 A General FGSL Analysis (skip)

When u_{it} is heteroskedastic and/or serially correlated, FGLS with $\hat{\mathbf{\Omega}} = N^{-1}\sum_{i=1}^{N}\hat{\mathbf{v}}_i\hat{\mathbf{v}}_i'$, $\hat{\mathbf{v}}_i = \mathbf{y_i} - \mathbf{X_i}\hat{\boldsymbol{\beta}}_{POLS}$ (Pooled OLS residual) works

10.4.4 Testing for the Presence of an Unobserved Effect

We can test the presence of unobserved effect c_i **by testing** $H_0 : \sigma_c^2 = 0$

Test 1(AR(1) test): t-test of the coefficient on AR(1) model of POLS residual

Test 2（Breusch and Pagan 1980）: Using asymptotic normality of $\sqrt{N}\hat{\sigma}_c^2$**, test** $H_0 : \sigma_c^2 = 0$

Test statistic: $\frac{\sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\hat{v}_{it}\hat{v}_{is}}{\left[\sum_{i=1}^{N}\left(\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\hat{v}_{it}\hat{v}_{is}\right)^2\right]^{1/2}}$, $\hat{\mathbf{v}}_i = \mathbf{y_i} - \mathbf{X_i}\hat{\boldsymbol{\beta}}_{POLS}$ **(Pooled OLS residual)**

asymptotically standard normal

10.5 Fixed Effects Methods

The unobserved effect model (UEM): $y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}$, $t = 1, \cdots, T$

Basic Idea: After eliminating unobserved effects by demean, and estimate by OLS

<u>**Assumption FE.1**</u>: $E(u_{it} \mid \mathbf{x}_i, c_i) = 0$, $t = 1, \cdots, T$ (strict exogeneity)

FE is more robust than RE because it allows any correlation between c and x

However, we cannot estimate the partial effects of variables constant across time (ex., race, sex, etc.)

<u>**Estimation: fixed effects transformation**</u>

(1) Construct time mean for each i $\bar{y}_i = \bar{\mathbf{x}}_i\boldsymbol{\beta} + c_i + \bar{u}_i$, $\bar{y}_i = T^{-1}\sum_{t=1}^{T} y_{it}$

(2) Demean: $y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\boldsymbol{\beta} + (u_{it} - \bar{u}_i)$ （note that c has disappears!）

Write it $\ddot{y}_{it} = \ddot{\mathbf{x}}_{it}\boldsymbol{\beta} + \ddot{u}_{it}$

(3) **Under AFE.1, POLS of this equation is consistent** (within estimator)

For the FE estimator to be consistent, we need $E(\ddot{\mathbf{x}}_{it}'\ddot{u}_{it}) = \mathbf{0}$, $t = 1, \cdots, T$. **This is satisfied under AFE.1 (strict exogeneity)**

Note : If we only assume contemporaneous exogeneity (not AFE1), Pooled OLS is no longer consistent!

Assumption for well-behaved asymptotic property

<u>**Assumption FE.2**</u>: $rank\left(\sum_{t=1}^{T} E(\ddot{\mathbf{x}}_{it}'\ddot{\mathbf{x}}_{it})\right) = K$, (rank condition)

If the model contains variables across time, this assumption is not satisfied. We need to exclude such variables form the model before FE estimation

Under AFE1 and 2, FE estimator is unbiased conditional on X.

Between estimator: OLS estimator of $\bar{y}_i = \bar{\mathbf{x}}_i\boldsymbol{\beta} + c_i + \bar{u}_i$ (estimator using cross-sectional variation only)

10.5.2 Asymptotic Inference with Fixed Effects

<u>**Assumption FE.3**</u>: $E(\mathbf{u}_i\mathbf{u}_i' \mid \mathbf{x}_i, c_i) = \sigma_u^2\mathbf{I}_T$, ($u_{it}$ has constant variance across t and are serially uncorrelated)

Under AFE3, FE estimator is efficient.

Stack the demeaned equation for each i: $\ddot{\mathbf{y}}_i = \ddot{\mathbf{X}}_i\boldsymbol{\beta} + \ddot{\mathbf{u}}_i$, $\ddot{\mathbf{y}}_i = (\ddot{y}_{i1}, \cdots, \ddot{y}_{iT})'$

$\hat{\beta}_{FE} = \left(\sum_{i=1}^{N}\ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i\right)^{-1}\left(\sum_{i=1}^{N}\ddot{\mathbf{X}}_i'\ddot{\mathbf{y}}_i\right)$ Under AFE.3, $\sqrt{N}(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}) \sim Normal(\mathbf{0}, \sigma_u^2[E(\ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i)]^{-1})$

$A\hat{\mathrm{var}}(\hat{\boldsymbol{\beta}}_{FE}) = \hat{\sigma}_u^2\left(\sum_{i=1}^{N}\ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i\right)^{-1}$, $\hat{\sigma}_u^2 = \frac{1}{N(T-1)-K}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{u}_{it}^2$, $\hat{u}_{it} = \ddot{y}_{it} - \ddot{\mathbf{x}}_{it}\hat{\boldsymbol{\beta}}_{FE}$ (FE residuals)

10.5.3 The Dummy Variable Regression (skip)

It is possible to estimate c_i by including all individual dummies in the model:

y_i on $d1_i,\ d2_i,\ \ldots,\ dN_i,\ \mathbf{x}_{it}$

We obtain the same estimates of betas from the estimation of the equation. Moreover, the individual specific effect can be estimated as coefficients on each dummy variable: $\hat{c}_i = \bar{y}_i - \bar{\mathbf{x}}_i\hat{\boldsymbol{\beta}}_{FE}$

10.5.4 Serial Correlation and the Robust Variance Matrix Estimator

Without AFE.3, we cannot use $A\,\hat{\mathrm{var}}(\hat{\boldsymbol{\beta}}_{FE}) = \hat{\sigma}_u^2\left(\sum_{i=1}^{N}\ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i\right)^{-1}$ **for inference.**

Firstly test u_{it}**'s serial correlation, and if it is detected, use robust variance matrix**

<u>**Serial correlation test 1**</u>

Construct $\hat{u}_{iT}$ **(fixed effect residuals). Regress** $\hat{u}_{iT}$ **on** $\hat{u}_{i,T-1}$ **and test for the coefficient on** $\hat{u}_{i,T-1}$**(**$\hat{\delta}$**),** $H_0 : \hat{\delta} = -1/(T-1)$ **(no correlation)**

Note : for $\delta = Cov(\ddot{u}_{iT}, \ddot{u}_{i,T-1})$**,** $\hat{u}_{iT}$**: fixed effect residuals is data-counterpart of** $\ddot{u}_{iT}$**.**

If $\hat{u}_{iT}$ **and** $\hat{u}_{i,T-1}$ **are uncorrelated,** $\delta = -1/(T-1)$ （t-test under **AFE1,2,3**）

<u>**Serial correlation test 2**</u>

Run pooled OLS $\hat{u}_{it}$ on $\hat{u}_{i,t-1}$ for $t = 3,\cdots,T; i = 1,\cdots,N$ and use the fully robust standard error for pooled OLS to test $H_0 : \hat{\delta} = -1/(T-1)$.

If serial correlation is detected, use

$$A\,\hat{\mathrm{var}}(\hat{\boldsymbol{\beta}}_{FE}) = \left(\ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i\right)^{-1}\left(\sum_{i=1}^{N}\ddot{\mathbf{X}}_i'\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i'\ddot{\mathbf{X}}_i\right)\left(\ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i\right)^{-1}$$

10.5.5 Fixed Effects GLS (skip)

Without AFE.3, we may estimate by GLS under a weaker assumption instead of calculating robust variance matrix. In such estimation, we need to drop any one cross-section

Assumption FEGLS.3: $E(\mathbf{u}_i\mathbf{u}_i' | \mathbf{x}_i, c_i) = \boldsymbol{\Lambda}$, a positive definite matrix.

10.5.6 Using Fixed Effects Estimation for Policy Analysis (skip)

Even if policy variable is correlated to persistent component of the error term($\bar{v}_i$), FE is consistent

10.6 First Differencing Methods

<u>**Assumption FD.1**</u>: $E(u_{it} \mid \mathbf{x}_i, c_i) = 0$, $t = 1, \cdots, T$ (strict exogeneity) same as AFE.1

<u>**Estimation : first differencing transformation**</u>

$\Delta y_{it} = \Delta \mathbf{x}_{it}\boldsymbol{\beta} + \Delta u_{it}$, $\Delta y_{it} = y_{it} - y_{i,t-1}, \Delta \mathbf{x}_{it} = \mathbf{x}_{it} - \mathbf{x}_{i,t-1}, \Delta u_{it} = u_{it} - u_{i,t-1}$

Under AFD.1, Pooled OLS is consistent (because $E(\Delta \mathbf{x}_{it}' \Delta u_{it}) = 0$**,** $t = 2, \cdots T$**)**

<u>**Assumption FD.2**</u>: $rank\left(\sum_{t=2}^{T} E(\Delta \mathbf{x}_{it}' \Delta \mathbf{x}_{it})\right) = K$, (rank condition)

Time-invariant variables must be dropped from the equation before FD estimation

10.6.2 Robust Variance Matrix: $A\,\hat{\text{var}}(\hat{\boldsymbol{\beta}}_{FD}) = (\Delta \mathbf{X}_i' \Delta \mathbf{X}_i)^{-1}\left(\sum_{i=1}^{N} \Delta \mathbf{X}_i' \hat{\mathbf{e}}_i \hat{\mathbf{e}}_i' \Delta \mathbf{X}_i\right)(\Delta \mathbf{X}_i' \Delta \mathbf{X}_i)^{-1}$

10.6.3 Testing for Serial Correlation: Estimate $\hat{e}_{it} = \hat{\rho}_1 \hat{e}_{i,t-1} + error_{it}$ and t-test $H_0 : \hat{\rho}_1 = 0$

10.6.4 Policy Analysis Using First Differencing

FD estimator is easier than FE estimator. However, Under AFE.1, 2, and 3 (homoskedasticity and no serial correlation), FE is more efficient than FD

Assumption that makes FD more efficient than FE

<u>**Assumption FD.3**</u>: $E(\mathbf{e}_i \mathbf{e}_i' \mid \mathbf{x}_i, c_i) = \sigma_e^2 \mathbf{I}_{T-1}$, where $\mathbf{e}_i = (e_{i2}, \cdots, e_{iT})'$ and $e_{it} = \Delta u_{it}$

(u_{it} is random walk)

Under AFD.1, 2, and 3, FD is efficient (among the class of estimators with strict exogeneity assumption)

$A\,\hat{\text{var}}(\hat{\boldsymbol{\beta}}_{FD}) = \hat{\sigma}_e^2 \left(\sum_{i=1}^{N} \Delta \mathbf{X}_i' \Delta \mathbf{X}_i\right)^{-1}$, $\hat{\sigma}_u^2 = \frac{1}{N(T-1)-K} \sum_{i=1}^{N} \sum_{t=2}^{T} \hat{e}_{it}^2$, $\hat{e}_{it} = \Delta y_{it} - \Delta \mathbf{x}_{it} \hat{\boldsymbol{\beta}}_{FD}$ (FD residuals)

Tests using above residuals are valid

10.7 Comparison of Estimators

10.7.1 Fixed Effects versus First Differencing

(1) When T=2, FE = FD

(2) When T>2, FE is more efficient than FD under AFE.3 (homoskedasticity and no serial correlation), while FD is more efficient than FD under AFD.3 (random walk).

Testing for strict exogeneity(skip)

Test with FD

$\Delta y_{it} = \Delta\mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{w}_{it}\boldsymbol{\gamma} + \Delta u_{it}$, $\mathbf{w}_{it}$: subset of $\mathbf{x}_{it}$ (excluding time dummies)

Under strict exogeneity, $\mathbf{x}_{it}$ should be insignificant. After estimating it by pooled OLS, test $H_0:\boldsymbol{\gamma}=\mathbf{0}$ by Wald（with robust-variance matrix if needed）(under AFD.1.-3, F-test is also valid)

Test with FE

$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{w}_{i,t+1}\boldsymbol{\delta} + c_i + u_{it}$, $\mathbf{w}_t$: subset of $\mathbf{x}_{it}$ (excluding time dummies)

Under strict exogeneity, $\mathbf{x}_{i,t+1}$ should be insignificant. Test $H_0:\boldsymbol{\delta}=\mathbf{0}$

10.7.2 The Relationship between the Random Effects and Fixed Effects Estimators

If $\mathbf{x}_{it}$ has small variation across time, we need to use RE

RE is equivalent to the Pooled OLS of the following equation(<u>quasi-demean</u>)

$y_{it} - \lambda\bar{y}_i = (\mathbf{x}_{it} - \lambda\bar{\mathbf{x}}_i)\boldsymbol{\beta} + (u_{it} - \lambda\bar{u}_i)$ where $\lambda = 1-[\sigma_u^2/(\sigma_u^2+T\sigma_c^2)]^{1/2}$

（Use $\hat{\lambda} = 1-[\hat{\sigma}_u^2/(\hat{\sigma}_u^2+T\hat{\sigma}_c^2)]^{1/2}$ for feasible RE）

Since $\lambda = 1-[1/(1+T(\sigma_c^2/\sigma_u^2)]^{1/2}$, when $T\to\infty$ or $(\sigma_c^2/\sigma_u^2)\to\infty$, $\lambda\to 1$ (same as FE). When $(\sigma_c^2/\sigma_u^2)\to 0$, $\lambda\to 0$ (same as Pooled OLS).

(FE allows any relationship between c and x. POLS assumes common individual effect c. RE is in the middle of the two because it assumes only orthogonality between c and x

10.7.3 The Hausman Test Comparing the RE and FE Estimators

<u>RE vs. FE? (are c and x uncorrelated?)</u>

H0: c and x are uncorrelated, H1: c and x are correlated

Under H0, both RE and FE are consistent. Moreover, under H0, RE is more efficient than FE

Under H1, only FE is consistent (RE is inconsistent)

Hausman Test: Under ARE1-3, asymptotically

$$H = \left(\hat{\delta}_{FE} - \hat{\delta}_{RE}\right)\left[A\,\hat{\mathrm{var}}(\hat{\delta}_{FE}) - A\,\hat{\mathrm{var}}(\hat{\delta}_{RE})\right]^{-1}\left(\hat{\delta}_{FE} - \hat{\delta}_{RE}\right) \sim \chi^2_M$$

(M: number of parameters to be estimated. δ_{FE}:Mx1.)

<u>Basic Idea</u>: When H1 is correct, RE estimator becomes very different from FE estimator because RE is inconsistent while FE is consistent. Hence , H becomes large when H1 is correct and reject H0

When H0 is true, since RE is more efficient than FE, variance of RE becomes “smaller” than that of FE, and therefore H becomes small and we do not reject H0

In Stata, command “hausman” is available. After estimation by RE and FE, it test Hausman test statistics

PS8

I. Reproduce the estimation results in Example 10.4 with Stata (hand in both the output and the DO files)

II. 10.1

III. 10.2

IV. Reproduce the estimation results in Example 10.5 (up to p.272) with Stata (hand in both the output and the DO files)

V. 10.7

VI. 10.10